



**QUEEN'S
UNIVERSITY
BELFAST**

Serious Games: Quality Characteristics Evaluation Framework and Case Study

Abdellatif, A. J., McCollum, B., & McMullan, P. (2018). Serious Games: Quality Characteristics Evaluation Framework and Case Study. In *2018 IEEE Integrated STEM Education Conference (ISEC): Proceedings* (pp. 112-119). Institute of Electrical and Electronics Engineers Inc.. <https://doi.org/10.1109/ISECon.2018.8340460>

Published in:

2018 IEEE Integrated STEM Education Conference (ISEC): Proceedings

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

© 2018 IEEE.

This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Serious Games: Quality Characteristics Evaluation Framework and Case Study

Abdelbaset Jamal Abdellatif, Barry McCollum, and Paul McMullan
Queen's University Belfast, aabdellatif01, b.mccollum, p.p.mcmullan@qub.ac.uk

Abstract – The use of serious games in teaching and training is increasing; however, there is a lack of suitable evaluation frameworks to evaluate the different quality characteristics in the serious games. This study highlights previous evaluation frameworks and then emphasizes different quality characteristics that have been used in evaluating serious games. The study divided the discussed quality characteristics into primary and secondary characteristics based on their use in the literature. The study proposes a framework to evaluate several dimensions of serious games by choosing and combining appropriate quality characteristics. Robocode a programming serious game was used as a case study in which the framework has been applied to it where fifteen students in Queen's University Belfast played the game and evaluated different quality characteristics based on the proposed framework. The results showed that Robocode overall evaluation is good; however, the framework recommends changes to be applied to the game to increase the game understandability to be played by the users without the need for supervision or tutors.

Index Terms - Evaluation framework, Quality characteristics, Robocode, Serious games, simulation software.

INTRODUCTION

Educational associations monopolize in information systems to obtain benefits such as increasing the accessibility of education, improving self-efficacy, cost-effectiveness, learner flexibility and interactivity [1]. Part of the investment is directed into serious games (SG) which are defined as “games that do not have entertainment, enjoyment or fun as their primary purpose” [2], instead, they offer educational content to users in an enjoyable way by simulating scenarios which promote learning. The positive relationship between using SG and students learning is currently perceived as a fact where considerable research findings have demonstrated the positive impact of SG on students learning outcomes and experience [3]-[11].

Evaluating SG and simulation software plays a crucial role in using such methods in learning. Interviews were conducted on teachers of programming modules in Queens University Belfast and it showed that there has been no use of SG in teaching computer programming due to the amount of time required to test the game before deploying and using

it on students. Yet, the teachers showed a willingness to use the SG in teaching computer programming if its effectiveness and suitability was proved.

Since limited games were used for educational purposes, the evaluation of SG has depended on studies of video games [12, 13]. However, evaluating SG is different than evaluating leisure games, and where there are several frameworks to evaluate leisure games, De Freitas, [14] stated that there is a shortage of useful frameworks for evaluating simulation software which resulted in a significant barrier for using them in teaching. Thus, [14, 15] specified the need for new frameworks and methods to evaluate SG.

The use of SG for supporting teaching in university and school levels without any use of a prior proper evaluation for the SG can limit the effectiveness and the learning outcomes. Where the use of an appropriate evaluation framework can provide guidelines and improvements to achieve the maximum benefits of SG.

PREVIOUS EVALUATION FRAMEWORKS

De Freitas and Oliver [16] developed a four-dimensional framework for helping tutors to evaluate the potential of SG and thus helping them in selecting a SG that matches their needs; further, the authors proposed that it can be used by game developers to design SG that match the needs of the educational program. The dimensions are context, learner specification, mode of representation and pedagogic considerations. The study emphasized the need to use the four dimensions together and not separately to illustrate the significance of the dimensions and how they relate to each other to support the learner experience. The context dimension focuses on where is the learning taking place in school, university, home or a combination of some, it also focuses on the need for specific resources and technical support and if it interferes with the context for learning. The learner specification dimension focuses on the learner attributes like age, background and learning styles. The mode of representation dimension focuses on what does the game represent, the realism of the game, the interactivity and the level of fidelity. The fourth dimension focuses on the process of learning in terms of the used pedagogic models, the curricular objectives, the learning outcomes and activities. However, Robertson and Howells, [17] stated that the four-dimensional framework requires the tutors to have a good background in computer games and furthermore it

doesn't sufficiently assist tutors in classifying the SG that match their needs.

Similarly, Dondi, *et al.* [18] presented a framework from two European projects for supporting teachers, trainers and professionals for the selection of the suitable SG that matches their needs and the changing educational requirements, and thus support the personal, social, cultural and cognitive development of their students.

Ssemugabi and Villiers [19] developed a framework for evaluating web-based learning applications, the main category in the framework was developed based on [20] heuristics, the other two categories covered the website-specific criteria for educational websites and learner-centred instructional design, grounded in learning theory, aiming for effective learning. Heuristic evaluation (HE) is an inspection technique where a set of experts evaluate whether a user interface conforms to defined usability principles, called heuristics [21]-[23], also HE is used to identifying specific usability problems in a system. HE can be applied to different quality characteristics other than usability like an evaluation of the games playability [24] and game design [25]. HE is useful however, it can be difficult, time-consuming and expensive [26]; moreover, it can lead to ignoring any usability problems that are not covered by the listed heuristics.

Xu, *et al.* [27] developed a SG framework assessment named Serious Game Stakeholder Experience Assessment Method (SGSEAM) that assesses SG frameworks from the major stakeholders' experience perspective. The goal of this framework is to identify the strengths and shortcomings of a SG framework for various stakeholders that are involved in the SG lifecycle including players, system admins, game designers, game managers, community partners and funding organizations. There are a set of questions and other assessment approaches proposed for each stakeholder which include the collection of quantitative and qualitative data. Using this framework can be helpful for identifying the usefulness of a SG and its suitability to be used for a party and highlighting the potential improvements that can be applied for each stakeholder; however, applying this framework on a large SG that involves multiple stakeholders can be difficult, expensive and time-consuming since for each stakeholder a data must be collected and analysed.

Schumann, *et al.* [28] construed and illustrated the four levels that formed Kirkpatrick, [29] framework for evaluating and analysing the results of training and educational programs; the four levels are reaction, learning, behaviour and results. The framework was developed to explore the effectiveness of the simulation software from four diverse perspectives. Farjad, [30] used Kirkpatrick model to evaluate the effectiveness of a training course at a university level. Betas, [31] referred the popularity of using Kirkpatrick model to provide a clear system or language and information needed to assess training programs and simplifying the complex evaluation process. On the other hand, Betas, [31] highlighted some limitations to this model

like incompleteness to the model due to not considering individual or contextual influences, the assumption of causal linkages, for example, the model assumes that positive reactions lead to greater learning that delivers greater transfer and thus positive results. Another limitation is that the model assumes that each level provides more informative data than the last level which produced a perception for evaluator that collecting level four data which is results will offer the most useful information. Furthermore, this model was designed for evaluating training programs and implementing it for evaluating SG in teaching will ignore multiple quality characteristics.

QUALITY CHARACTERISTICS

The literature in evaluating SG is diverse and different quality characteristics can be used to evaluate several aspects of SG, A noticeable effort has been made by [32] who analyzed the literature and summarized quality characteristics that have been used to evaluate SG into 18 characteristics which are game design, user's satisfaction, usability, usefulness, understandability, motivation, performance, playability, pedagogical aspects, learning outcomes, engagement, user's experience, efficacy, social impact, cognitive behavior, enjoyment, acceptance and user interface.

These quality characteristics will be explored in the following section to assess the relevance of each characteristic to this research program.

I. Game Design

Which refers to how a SG is designed and the appealing and artistic visual design. Deterding, *et al.* [33] stated that there are 5 levels of game design elements which are game interface design patterns, game design patterns and mechanics, game design principles and heuristics, game models and game design methods. Muratet, *et al.* [34] considered game design and its appealing look as part of user's satisfaction. Game design has been measured and used in different studies like [34, 35].

II. User's Satisfaction

Which does not have a unified definition but rather depending on context and population subgroups [36]. Several factors affect user satisfaction like perceived usefulness, perceived ease of use and perceived flexibility [37]-[39]. [40, 41] highlighted other factors like age, gender, users' initial computer skills and users' initial knowledge about e-learning technology. Moreover, [42, 43] underlined further factors like motivation and attitude towards technology. Mortara, *et al.* [44] developed a six dimensions framework to assess e-learning satisfaction which are learner dimension, instructor dimension, course dimension, technology dimension, design dimension and environmental dimension. User's satisfaction has been measured and used in several studies like [45, 46].

III. Usability

It has been defined by the international standard ISO 9241-11 as the “extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [47]. According to the definition, usability is a construct consisting of three dimensions: effectiveness, efficiency and satisfaction. Dumas and Redish [48] defined the usability as “the people who use the product can do so quickly and easily to accomplish their own tasks” further, the book identified four points that form the definition which are focusing on users, people use products to be productive, users are busy trying to accomplish tasks and users decide when a product is easy to use. Similarly, Nielsen, [49] stated the usability is associated with five attributes which are learnability, efficiency, memorability, errors, and satisfaction. Many studies measured the usability characteristic in SG like [34], [50]-[53].

IV. Usefulness

It was defined as “The degree to which a person believes that using a particular system would enhance his or her job performance” [54]. Alsabawy, *et al.* [55] stated that perceived usefulness is the main measure to assess the acceptance and success of e-learning systems; moreover, Sela and Sivan, [56] listed usefulness as the first ‘Must-Have’ success factor in e-learning. In the context of SG, usefulness refers to the efficiency and suitability of the SG for the designated audience, it has been measured as a quality characteristic in various studies like [57, 58].

V. Understandability

It is defined in the ISO 9126 document as “the capability of the component to enable the user to understand whether it suitable, and how it can be used for particular tasks and conditions of use” [59], it’s considered part of the usability characteristic along with learnability, operability, attractiveness and usability compliance. Another definition for understandability was listed by Boehm, *et al.* [60] where software understandability was defined as a characteristic of software quality which means ease of understanding software systems. Understandability is a crucial characteristic where if a SG or a software is not understandable it will not achieve its educational goals. This characteristic has been explored and measured in studies like [61]-[63].

VI. Motivation

Which refers to the ability of a used SG to impact and affect the user’s motivation where it encourages the user to use the SG. According to Wrzesien, *et al.* [64], SG boost the user’s intrinsic motivation where it leads the users to curiosity and desire for challenge. Motivation is considered an important characteristic in which several theories like [65] highlighted the potential of SG to positively impact intrinsic motivation and furthermore the study stated that probably the primary purpose for using SG is their assumed motivational appeal.

Moreover, Girard, *et al.* [66] analysis of previous studies on the effectiveness of SG explored the motivational factor as a primary key for an effective use of a SG. The motivation characteristic has been used and measured by numerous studies like [67]-[70].

VII. Performance

Which can be described as the functioning and the effectiveness of a SG. The performance characteristic has been measured in studies by rating the functionalities of the SG [52]; also, it has been used to evaluate a SG in the development process to test the components of the game [71].

VIII. Playability

Which means is the SG playable, it is a crucial characteristic as if the SG is not playable then it will fail to accomplish its designated purpose. Olsen, *et al.* [53] highlighted the importance of playability characteristic in SG. Playability focuses on the complete functionality linked with the integration of numerous usable tools tolerating for effective and pleasant interaction with the game. Heuristic evaluation has been developed to evaluate the playability of games [24] and it has been used to evaluate SG [72]. Manero, *et al.* [71] evaluated playability as a main quality characteristic to evaluate the effectiveness of the used SG.

IX. Pedagogical Aspects

Which refers to the educational content provided by a SG. It is vital that the educational content of a SG is well adjusted and calibrated, as Quinn, [73] argued that SG must be designed appropriately to create a harmony between the gameplay and the learning objectives. [71, 74] evaluated the pedagogical aspects in their evaluation of the SG.

X. Learning Outcomes

It is considered the most important characteristic since its presence convert the game into a serious and educational game. Learning outcomes refers to the knowledge gained by the users after interacting with the SG. Most of the studies evaluate the learning outcomes of a SG first and then consider other dimensions or characteristics. In computer programming SG, many studies evaluated the learning outcomes and used qualitative and/or quantitative methods and data in their analysis [4]-[10], [67]. Some studies divided learning outcomes into categories like Wouters, *et al.* [75] where they divided learning outcomes into four categories cognitive, motor skills, affective and communicative.

XI. Engagement

Which refers to engage users in using and playing a SG; engagement is complex as stated by Dele-Ajayi, *et al.* [76] where the study considered motivation as the root of engagement. Pourabdollahiana, *et al.* [77] used a framework for evaluating the engagement characteristic based on two studies [78, 79], the framework consisted of five factors that

affect engagement which are immersion, control, challenge, purpose and interest. Several studies evaluated the engagement characteristic in SG like [80, 81].

XII. User's Experience

It is defined in the ISO 9241 document as “A person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service” [82]. In Wangenhei, *et al.* [67] evaluation model to evaluate a SG, the study recommended three factors to assess a used SG which are motivation, learning and user experience, under user experience factor the study listed five aspects which are immersion, challenge, competence, fun and social interaction. A considerable number of studies evaluated user's experience characteristic like [67], [83]-[85].

XIII. Efficacy

Which was defined as “the power or ability of the game to improve participants' knowledge” [86]. Evaluating the efficacy of a SG involves evaluating the engagement, attention, involvement, enjoyment, difficulties, and time to complete the game [87]-[89]. Several studies evaluated the effectiveness of SG like [10, 64, 86, 90, 91].

XIV. Social Impact

It can be defined as the impact or the influence that SG create in the users or contribute to it. Swain, [92] listed 8 practices for designing and developing SG to affect social change which are define intended outcomes, integrate subject matter experts, partner with like-minded organizations, build a sustainable community, embrace “wicked problems”, maintain journalistic integrity, measure transference of knowledge and make it fun. Hensman, [93] stated an advantage of the used SG “as a platform for students to showcase their work in a socially interactive environment”. Measuring social impact characteristic of SG is limited in the literature and one example of measuring this characteristic is [94].

XV. Cognitive Behaviour

It can be identified as the ability of SG to create effects and changes in user's cognitive behaviour. Hauge and Riedel, [74] stated that cognitive analysis is a beneficial tool that allows game developers to well tune a SG. Cognitive behaviour analysis is limited in the literature, few studies like [95, 96] measured this quality characteristic.

XVI. Enjoyment

It can be defined as the joy and fun that the SG produces to the users. Ricci, [97] stated that SG increase the user interest because users enjoy this approach to learn. In Hensman, [93] study about a SG called Robocode, the study questioned the enjoyable aspects of the game and the answers were its competitive nature and helping to learn computer programming. Enjoyment is considered as a key factor in games in general where previous studies showed that enjoyment of playing is one of the main motivators for

playing computer games [98, 99]. Assorted studies explored the enjoyment characteristics like [10, 52, 64].

XVII. Acceptance

Which refers to the SG being accepted and received by the user. According to Alsabawy, *et al.* [55], the acceptance characteristic can be assessed by measuring the usefulness of a system, where acceptance is linked to perceived usefulness and perceived ease of use [54]. Acceptance is considered as an important characteristic since if the user didn't accept a game and used, he/she will not engage effectively and thus will not receive the specified knowledge. Brom, *et al.* [100] identified five key factors that led to the positive acceptance of the used game which are intelligibility, social role-playing, grounding in real data, story-telling and support for teachers. Technology Acceptance Model (TAM) has been used to predict the user acceptance of new technology [54]. Yusoff, *et al.* [101] proposed a model for designing SG and using TAM for validating the acceptance of the game. Few studies considered measuring the acceptance characteristic of SG like [100, 102, 103].

XVIII. User Interface

Which refers to the interaction that occurs between a user and SG. Lanyi, *et al.* [104] stated that it is essential to design a user interfaces for maximum accessibility and usability. Mikovec, *et al.* [105] investigated the designing of a user interface for SG and the need for developing a user interface that is usable and understandable; furthermore, the study stated that SG are barely successful without an appropriate and well-designed user interface. Deterding, *et al.* [33] stated that game interface is one level of the elements of game design. Few studies measured user interface characteristic like [104].

FRAMEWORK DEVELOPMENT

In this section, the quality characteristics are divided into primary and secondary based on their importance and then the proposed framework is explained.

I. Primary Quality Characteristics

Refers to the characteristics that their absence will prevent SG from delivering its educational content to a designated audience effectively. Learning outcomes evaluation is widely conducted and considered as a primary characteristic, along with pedagogical aspects characteristic they form the difference between video games and SG where their occurrence makes games have educational content and purpose. For evaluating these characteristics an experiment must be conducted in which a control group and an experiment group should take a place to measure if a SG has succeeded in delivering its educational content by comparing the results of the two groups. Evaluating the learning outcomes of a SG without conducting experiments could only be done to evaluate the potential of a SG and its suitability for the selected audience only. Similarly, the user

satisfaction characteristic must be measured by running an experiment and collecting feedback from users, this characteristic depends on many other quality characteristics like game design and interface. Engagement, motivation and user experience are primary characteristics since their appearance pushes the user to engage effectively with a game and with other users. Another primary characteristic is usability which is considered one of the most measured quality characteristics for SG. Understandability characteristic refers to the clarity of a game and measuring this characteristic is limited in the literature, yet its absence prevents the SG from delivering its content and thus not achieving its purpose, for that understandability is considered a primary characteristic in this study.

II. Secondary Quality Characteristics

Refers to the characteristics that aren't crucial to the success of SG in delivering its educational content for the designated audience. Social impact and cognitive behaviour quality characteristics have limited usage in evaluation frameworks and can be found in evaluating SG that are meant for children and thus they are considered as secondary characteristics. Evaluating game design and user interface characteristics are limited and these two characteristics are linked to each other, they depend on two primary characteristics which are usability and user satisfaction and thus there is no need to measure them independently if the other primary characteristics are measured. Acceptance, usefulness, performance, efficacy and playability characteristics are part of other primary characteristics in which measuring learning outcomes and user's satisfaction will provide an overall view of these characteristics, but this study doesn't consider the need for those characteristics to be measured individually. Enjoyment characteristic relays on several secondary characteristics like playability and game design, also it is part of the motivation and engagement characteristics and can be measured as part of the other primary characteristics rather than being measured on its own.

III. The Framework

The framework was designed to measure the primary quality characteristics that don't require conducting an experiment that involves tests with students and can be applied in a brief period. The framework consists of five primary quality characteristics which are displayed in figure 1, and for each characteristic three questions were assigned to measure specific factors, below is the list of the characteristics and what are they evaluating. First, usability and it measures the learnability, usefulness, errors and ease of use. Second, motivation and it measures challenge, enjoyment and curiosity. Third, engagement and it measures purpose, interest and control. Fourth, user experience and it measures competence, social interaction and fun. Fifth, Understandability and it measures clarity and the need for supervision.

DISCUSSION AND RESULTS

Data has been collected from 16 final year business information technology students who were working on computing IT enterprise project in Computer Science school. They voluntarily participated by using and playing the case study game called Robocode. Robocode is an open source Java-based virtual robot game that is destined to teach object-oriented programming concepts.



FIGURE I
FIVE DIMENSIONS EVALUATION CRITERIA

The Robocode game consists of a robot-development tool and it simulates a virtual battlefield where robots can battle against each other. The player programs the robot commanding it how to perform and respond to events arising on the battlefield. Thus, Robocode forms a space for students and learners to learn and apply their knowledge in OOP in which it covers writing classes, reading, analysing and using existed code, event handling and message passing [106]. After playing the game, the students were asked to rate 15 evaluation factors on a scale from (1-10) in which every 3 statements covered a quality characteristic from the proposed framework. We have got 15 valid responses of which 10 are males and 5 are females and all the respondents are in the age group between 18 and 23.

The average was calculated for each quality characteristic out of 100 by combining the results for the three evaluation factors that represent the quality characteristic as shown in table 1.

TABLE I
QUALITY CHARACTERISTIC RATING

No	Quality Characteristic	Average out of 100
1	Usability	71.3%
2	Understandability	62.2%
3	Motivation	69.5%
4	Engagement	66%
5	User experience	66.4%

The usability characteristic achieved the highest rating from the students with 71.3% while the understandability scored the lowest with 62.2%. As shown in table 1, the

scores indicate that understandability is an issue in the use of SG in teaching. Understandability was measured by three factors which are:

- The game goals are clear and understood easily.
- The game offers a set of straightforward steps to be followed.
- The game can be played individually without the need for assistance.

Furthermore, a student added this comment “The main problem was that I had no idea what was going on when I first started it. A little bit of an explanation would help”. Understandability is a complex characteristic that affects other characteristics such as engagement and user experience. Thus, these results suggest the need for more emphasis on improving the understandability characteristic of SG. Based on the three understandability factors used in this study, the problem is understanding game goals and how to use the game itself. This issue for Robocode could be solved by two ways, first by assigning tutors to explain the game goals for the students and by showing them how to use the game; second by adding a tutorials option for the game that students can follow so they can grasp the required information that will allow them to understand the game and use it on their own.

CONCLUSION

Previous studies attempted to evaluate the use of SG and proposed frameworks such as the four-dimensional framework; however, these frameworks fall short in considering different quality characteristics for SG which would affect the opportunity to optimize the development of SG or the actual use of the SG on the designated audience.

This study proposed a new framework for evaluating the use of SG by considering several quality characteristics which are; usability, understandability, motivation, engagement and user experience. The case study results indicate that understandability is an issue and therefore recommends the use of tutors or adding tutorials to the game to enhance users understanding of the goals, concepts and procedures of the game.

Further, understandability might affect other quality characteristics such as engagement and user experience which would be a future research direction to test this relationship by adding tutorials to the SG and repeating the evaluation to validate the results.

Currently, there is an attempt to apply this framework on university year 1 students and school students that are taking programming courses; the successful implementation of this framework will add more validity to the obtained results and generalise the results to include university and school levels.

REFERENCES

- [1] Sinclair, P. M., Kable, A., Levett-Jones, T. and Booth, D., (2016). “The effectiveness of internet-based e-learning on clinician behaviour and patient outcomes: A systematic review”. *International Journal of Nursing Studies*, Vol. 57, pp. 70-81.
- [2] Michael, D. and Chen, S., (2006). “Serious Games: That Educate, Train, and Info”. *Thomson Course Technology*.
- [3] Al-Linjawi, A. A. and Al-Nuaim, A. H., (2010), “Using Alice to Teach Novice Programmers OOP Concepts”, *Journal of King Abdulaziz University-Science*, Vol. 22, No. 1, pp. 59-68.
- [4] Baker, A., Zhang, J. and Caldwell, R. E., (2012). “Reinforcing Array and Loop Concepts Through a Game-Like Module”. *The 17th International Conference on Computer Games*, pp. 175-179.
- [5] Corral, J., Balcells, A., Estévez, A., et al., (2014). “A game-based approach to the teaching of object-oriented programming languages”. *Computers & Education*, Vol. 73, pp. 83-92.
- [6] Eagle, M., and Barnes, T., (2008). “Wu’s Castle: Teaching Arrays and Loops in a Game”. *ACM SIGCSE Bulletin*, Vol. 40, pp. 245-249.
- [7] Galves, J., Guzman, E. and Conejo, R., (2009). “A blended E-learning experience in a course of object oriented programming fundamentals”. *Knowledge-Based Systems*, Vol. 22, pp. 279-286.
- [8] Hillyard, C., Angotti, R., Panitz, M., Sung, K., Nordlinger, J. and Goldstein, D., (2010). “Game-Themed Programming Assignments For Faculty: A Case Study”. In *Proceedings of the 41st ACM technical symposium on Computer science education*, pp. 270-274.
- [9] Liu, P., (2008). “Using open-source Robocode as a java programming assignment”. *SIGCSE Bulletin*, vol. 40, Issue. 4, pp. 63-67.
- [10] Malliarakis, C., Satratzemi, M. and Xinogalos, S., (2014). “CMX: Implementing an MMORPG for Learning Programming”. *Proceeding of the European Conference on Games-based Learning*, pp. 346-355.
- [11] Zhang, J., Caldwell, R.E. and Smith, E., (2013). “Learning the Concept of Java Inheritance in a Game”. *The 18th International Conference on Computer Games (CGAMES)*, pp. 212-216.
- [12] Gee, J., (2003). “What video games have to teach us about learning and literacy”. *New York: Palgrave Macmillan*.
- [13] Squire, K., (2002). “Cultural framing of computer/video games”. *Game Studies*, Vol. 2. Available at: <http://www.gamestudies.org/0102/squire>.
- [14] De Freitas, S., (2004). “Learning through Play. Internal report”. *London: Learning and Skills Research Centre*.
- [15] De Freitas, S., (2005). “Review of the uptake and embedding of digital content”. *Internal report. Coventry: Becta*.
- [16] De Freitas, S. & Oliver, M., (2006). “How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?”. *Computers & Education*, Vol. 46, pp.249-264.
- [17] Robertson, J. & Howells, C., (2008). “Computer game design: Opportunities for successful learning”. *Computers & Education*, Vol. 50, Issue. 2, pp. 559-578.
- [18] Dondi, C. and Moretti, M., (2007). “A methodological proposal for learning games selection and quality assessment”. *British Journal of Educational Technology*, Vol. 38, No. 3, pp. 502-512.
- [19] Ssemugabi, S. and de Villiers, R., (2007). “A comparative study of two usability evaluation methods using a web-based e-learning application”. In *Proceedings of the 2007 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*, pp. 132-142.
- [20] Nielsen, J. and Mack, R. L., (1994). “Usability Inspection Methods”. *John Wiley & Sons, New York, NY*.
- [21] Dix, J., Finlay, E., Abowd, D. et al., (2004). “Human-Computer Interaction”. 3rd Ed. Harlow Essex: Pearson Education Limited.
- [22] Nielsen, J. 1992. Finding Usability Problems through Heuristic Evaluation. In: *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*: 373- 380. Monterey: ACM Press.
- [23] Preece, J., Rogers, Y. and Sharp, H., (2002). “Interaction Design: Beyond Human-Computer Interaction”. *New York: John Wiley & Sons*.
- [24] Desurvire, H., Caplan, M. and Toth, A. J., (2004). “Using Heuristics to Evaluate the Playability of Games”. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, pp. 1509-1512.
- [25] Song, S. and Lee, J., (2007). “Key factors of heuristic evaluation for game design: Towards massively multi-player online role-playing game”. *International Journal of Human-Computer Studies archive*, Vol. 65, Issue. 8, pp. 709-723.
- [26] Kjeldskov, J., Skov, B. and Stage, J., (2004). “Instant Data Analysis: Conducting Usability Evaluations in a Day”. *Proceedings of the third Nordic Conference on Human Computer Interaction*, pp. 233-240.

- [27] Xu, Y., Johnson, M. P., Moore, A. C., Brewer, S. R. and Takayama, J., (2013). "SGSEAM: assessing serious game frameworks from a stakeholder experience perspective". *First International Conference on Gameful Design, Research, and Applications*, pp. 75-78.
- [28] Schumann, P., Anderson, H. P., Scott, W. T. and Lawton, L., (2001). "A Framework For Evaluating Simulations As Educational Tools". *Developments in Business Simulation and Experiential Learning*, Vol. 28, pp. 215-220.
- [29] Kirkpatrick, D. L., (1998). "Evaluating Training Programs: The Four Levels (2nd Ed)". *San Francisco: BerrettKoehler*.
- [30] Farjad, S., (2012). "The Evaluation Effectiveness of training courses in University by Kirkpatrick Model (case study: Islamshahr university)". *Social and Behavioral Sciences* Vol. 46, pp. 2837-2841.
- [31] Betas, R., (2004), "A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence". *Evaluation and Program Planning*, Vol. 27, pp. 341-347.
- [32] Calderon, A. and Ruiz, M., (2015). "A systematic literature review on serious games evaluation: An application to software project management". *Computers & Education*, Vol. 87, pp. 396-422.
- [33] Deterding, S., Dixon, D., Khaled, R. and Nacke, L., (2011). "From game design elements to gamefulness: defining "gamification"". *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, pp. 9-15.
- [34] Muratet, M., Torguet, P., Viallet, F., et al., (2011)., "Experimental feedback on Prog&Play: a serious game for programming practice". *Computer Graphics Forum*, Vol. 30, pp. 61-73.
- [35] Cederholm, H., Hilborn, O., Lindley, C., et al., (2011). "The aiming game: using a game with biofeedback for training in emotion regulation". In *Proceedings of the 5th International Conference on Digital Research Association: Think Design Play*.
- [36] Doll, J. W., Deng, X., Raghunathan, S. T., et al., (2004). "The Meaning and Measurement of User Satisfaction: A Multigroup Invariance Analysis of the End-User Computing Satisfaction Instrument". *Journal of Management Information System*, Vol. 21, Issue. 1, pp. 227-262.
- [37] Arbaugh, J. B., (2000). "Virtual classroom characteristics and student satisfaction with internet-based MBA courses". *Journal of Management Education*, Vol. 24, Issue. 1, pp. 32-54.
- [38] Arbaugh, J. B., (2002). "Managing the on-line classroom: a study of technological and behavioral characteristics of web-based MBA courses". *Journal of High Technology Management Research*, Vol. 13, pp. 203-223.
- [39] Arbaugh, J. B., & Duray, R., (2002). "Technological and structural characteristics, student learning and satisfaction with web-based courses – An exploratory study of two on-line MBA programs". *Management Learning*, Vol. 33, Issue. 3, pp. 331-347.
- [40] Hong, K. S., (2002). "Relationships between students' and instructional variables with satisfaction and learning from a Web-based course". *Internet and Higher Education*, Vol. 5, pp. 267-281.
- [41] Thurmond, A., Wambach, K. and Connors, R., (2002). "Evaluation of student satisfaction: determining the impact of a web-based environment by controlling for student characteristics". *The American Journal of Distance Education*, Vol. 16, Issue. 3, pp. 169-189.
- [42] Kanuka, H., & Nocente, N., (2003). "Exploring the effects of personality type on perceived satisfaction with web-based learning in continuing professional development". *Decision Education*, Vol. 24, Issue. 2, pp. 227-245.
- [43] Piccoli, G., Ahmad, R., & Ives, B., (2001). "Web-based virtual learning environments: a research framework and a preliminary assessment of effectiveness in basic IT skill training". *MIS Quarterly*, Vol. 25, Issue. 4, pp. 401-426.
- [44] Sun, P., Tsai, R., Finger, G. et al., (2008). "What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction". *Computers & Education*, Vol. 50, Issue. 4, pp. 1183-1202.
- [45] Mortara, M., Catalano, E., Fiucci, G., et al., (2014). "Evaluating the effectiveness of serious games for cultural awareness: the Icura user study". *Lecture Notes in Computer Science*, pp.276-289.
- [46] Torrente, J., Moreno-Ger, P., Fernandez-Manjon, B. et al., (2009). "Game-like simulations for online adaptive learning: a case study". *Lecture Notes in Computer Science*, Vol. 5670, pp. 162-173.
- [47] ISO/IEC 9126:1998 Ergonomic requirements for office work with visual display terminals - Part 11: Guidance on usability. (1998).
- [48] Dumas, J. and Redish, J., (1999). "A Practical Guide to Usability Testing". *Intellect Books Exeter, UK*.
- [49] Nielsen, J., (1993). "Usability Engineering". *Morgan Kaufmann Publishers Inc. San Francisco, CA, USA*.
- [50] Rodríguez-Cerezo, D., Sarasa-Cabezuelo, A., Gomez-Albarran, M. and Sierra, J., (2014). "Serious games in tertiary education: a case study concerning the comprehension of basic concepts in computer language implementation courses". *Computers in Human Behavior*, Vol. 31, pp. 558-570.
- [51] Couceiro, M. R., Papastergiou, M., Kordaki, M. and Veloso, I. A., (2013). "Design and evaluation of a computer game for the learning of Information and Communication Technologies (ICT) concepts by physical education and sport science students". *Education and Information Technologies*, Vol. 18, Issue. 3, pp. 531-554.
- [52] Zhang, J., Caldwell, R.E. and Smith, E., (2013). "Learning the Concept of Java Inheritance in a Game". *The 18th International Conference on Computer Games (CGAMES)*, pp. 212-216.
- [53] Olsen, T., Procci, K. and Bowers, C., (2011). "Serious games usability testing: how to ensure proper usability, playability, and effectiveness". *Lecture Notes in Computer Science*, Vol. 6770, pp. 625-634.
- [54] Davis, D. F., (1989). "Perceived usefulness, perceived ease of use and user acceptance of information technology". *MIS Quarterly*, Vol. 13, Issue. 3, pp. 319-340.
- [55] Alsabawy, Y. A., Cater-Steel, A. and Soar, J., (2016). "Determinants of perceived usefulness of e-learning systems". *Computers in Human Behavior*, pp. 843-858.
- [56] Sela, E. and Sivan, Y. Y., (2009). "Enterprise e-learning success Factors: An analysis of practitioners' perspective". *Interdisciplinary Journal of E-learning and Learning Objects*, Vol. 5, pp. 335-343.
- [57] Costantino, F., Di Gravio, G., Shaban, A. et al., (2012). "A simulation based game approach for teaching operations management topics". In *Proceedings of the 2012 Winter Simulation Conference*, pp. 1-12.
- [58] Backlund, P., Taylor, A., Engstrom, H., et al., (2011). "Evaluation of usefulness of the Elinor Console for homebased stroke rehabilitation". In *Proceedings of the 3rd International Conference on Games and Virtual Worlds for Serious Applications*, pp. 98-103.
- [59] ISO/IEC 9126:2001 Software engineering - Product quality - Part 1: Quality model. International Organization for Standardization and International Electrotechnical Commission. (2001).
- [60] Boehm, W., Brown, R., Kaspar, H. et al., (1978). "Characteristics of Software Quality". *Amsterdam, North Holland*.
- [61] Buttussi, F. and Chittaro, L., (2010). "Smarter phones for healthier lifestyles: an adaptive fitness game". *IEEE Pervasive Computing*, Vol. 9, Issue. 4, pp. 51-57.
- [62] Jung, H., Kim, S. and Chung, C., (2004). "Measuring software product quality: a survey of ISO/IEC 9126". *IEEE Software*, Vol. 21, Issue. 5, pp. 88-92.
- [63] Uchida, S. and Shima, K., (2004). "An Experiment of Evaluating Software Understandability". *Systemics, Cybernetics and Informatics*, Vol. 2, Number. 5, pp. 7-11.
- [64] Wrzesien, M. and Alcaniz, M., (2010). "Learning in serious virtual worlds: evaluation of learning effectiveness and appeal to students in the E-Junior project". *Computers & Education*, Vol. 55, pp. 178-187.
- [65] Garriss, R., Ahlers, R. and Driskell, J. E., (2002). "Games, Motivation, and Learning: A Research and Practice Model", *Simulation and Gaming*, Vol. 33, pp. 441-467.
- [66] Girard, C., Ecalle, J. and Magnant, A., (2013). "Serious games as new educational tools: how effective are they? A meta-analysis of recent studies". *Journal of Computer Assisted Learning*, Vol. 29.
- [67] Wangenheim, G. C., Savi, R. and Borgatto, F. A., (2012). "DELIVER! – An educational game for teaching Earned Value Management in computing courses". *Information and Software Technology*, Vol. 54, pp. 286-298.
- [68] Miller L. M., Chang C.-I., Wang S., Beier M. E. and Klisch Y., (2011). "Learning and motivational impacts of a multimedia science game". *Computers & Education* Vol. 57, pp. 1425-1433.
- [69] Kebritchi M., Hirumi A. and Bai H., (2010). "The effects of modern mathematics computer games on mathematics achievement and class motivation". *Computers & Education* Vol. 55, pp. 427-443.

- [70] Papastergiou, M., (2009). "Digital game-based learning in high school computer science education: impact on educational effectiveness and student motivation". *Computers & Education Vol. 52*, pp. 1–12.
- [71] Manero, B., Fernandez-Vara, C. and Fernandez-Manjon, B., (2013). "E-Learning Takes the Stage: From La Dama Boba to a Serious Game". *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje, Vol. 8, Issue. 4*, pp. 179-204.
- [72] Petrasova, A., Cza, S., Chalmers, A., Farrer, J. and Wolke, D., (2010). "The playability evaluation of virtual baby feeding application". In *VS-Games 2010-Proceedings of the 2nd International Conference on Games and Virtual Worlds for Serious Applications*, pp. 95-100.
- [73] Quinn, N. C., (2005). "The Seven Step Program for eLearning Improvement". Available at: <https://quinnovation.com/EnhancedIDWP.pdf>.
- [74] Hauge, J. B. and Riedel, J. C., (2012). "Evaluation of simulation games for teaching engineering and manufacturing". *Procedia Computer Science, Vol. 15*, pp. 210-220.
- [75] Wouters, P., van der Spek, E. and van Oostendorp, H., (2009). "Current practices in serious game research: A review from a learning outcomes perspective". *Games-based learning advancements for multisensory human computer interfaces: Techniques and effective practices*, pp. 232-255.
- [76] Dele-Ajayi, O., Strachan, R., Sanderson, J. and Pickard, A., (2016). "Learning Mathematics Through Serious Games: An engagement framework". In *Frontiers in Education Conference (FIE)*.
- [77] Pourabdollahiana, B., Taischa, M. and Kergaa, E., (2012). "Serious Games in Manufacturing Education: Evaluation of Learners' Engagement". *Procedia Computer Science, Vol. 15*, pp. 256-265.
- [78] Csikszentmihalyi M., (1992). "Flow: the Psychology of Happiness". London: Random House.
- [79] Malone, T. W. and Lepper, M., (1987). "Making Learning fun: A Taxonomy of intrinsic motivations for learning". In *Aptitude, learning and instruction. Cognitive and affective process analysis*. Lawrence Erlbaum, Hillsdale, N.J, pp. 223-253.
- [80] Adamo-Villani, N., Haley-Hermiz, T. and Cutler, R., (2013). "Using a serious game approach to teach 'operator precedence' to introductory programming students". In *Proceedings of the 17th International Conference on Information Visualisation*, pp. 523-526.
- [81] Dunwell, I., Lameris, P., Star, K., et al., (2013). "Metycoon: a game-based approach to career guidance". In *VS-Games 2013- Proceedings of the 5th International Conference on Games and Virtual Worlds for Serious Applications*, pp. 1-6.
- [82] ISO 9241-210:2010 Ergonomics of human-system interaction—Part 210: Human-centered design for interactive systems. (2010).
- [83] Hou, H. and Li, M., (2014). "Evaluating multiple aspects of a digital educational problem-solving-based adventure game". *Computers in Human Behavior, Vol. 30*, pp. 29-38.
- [84] Law, L. E. and Sun, X., (2012). "Evaluating user experience of adaptive digital educational games with Activity Theory". *International Journal of Human-Computer Studies, Vol. 70 Issue. 7*, pp. 478-497.
- [85] Pavlas, D., Jentsch, F., Salas, E., Fiore, S. M. and Sims, V., (2012). "The Play Experience Scale: development and validation of a measure of play". *The Journal of the Human Factors and Ergonomics Society, Vol. 54, Issue. 2*, pp. 214-225.
- [86] Ingadottir, B., Blondal, K., Thue, D., et al., (2017). "Development, Usability, and Efficacy of a Serious Game to Help Patients Learn About Pain Management After Surgery: An Evaluation Study". *JMIR Serious Games, Vol. 5, No. 2*.
- [87] Faria, A. J., Hutchinson, D., Wellington, W. J. and Gold, S., (2009). "Developments in business gaming: a review of the past 40 years". *Simulation and Gaming, Vol. 40, Issue. 4*, pp. 464–487.
- [88] Hartevelt, C., Guimarães, R., Mayer, S., et al., (2010). "Balancing play, meaning and reality: the design philosophy of LEVEE PATROLLER". *Simulation & Gaming, Vol. 41*, pp. 316–340.
- [89] Thompson, D., Baranowski, T., Buday, R., et al., (2010). "Serious video games for health: how behavioral science guided the development of a serious video game". *Simulation Gaming, Vol. 41, Issue. 4*, pp. 587–606.
- [90] Guillén-Nieto, V. and Aleson-Carbonell, M., (2012). "Serious games and learning effectiveness: The case of It's a Deal!". *Computers & Education, Vol. 58*, pp. 435-488.
- [91] Mayo, M. J., (2007). "Games for science and engineering education". *Communications of the ACM, Vol. 50, Issue. 7*, pp. 31–35.
- [92] Swain, C., (2007). "Designing Games to Effect Social Change". *Proceedings of DiGRA 2007 Conference*.
- [93] Hensman, A., (2007). "Evaluation of Robocode as a Teaching Tool for Computer Programming". *Proceedings of the National Digital Learning Repository Symposium, Trinity College Dublin, Ireland*.
- [94] Rebolledo-Mendez, G., Avramides, K., de Freitas, S. and Memarzia, K., (2009). "Societal impact of a serious game on raising public awareness: the case of FloodSim". In *Sandbox 2009-Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games*, pp. 15-22.
- [95] Manera, V., Petit, D., Derreumaux, A. et al., (2015). "Kitchen and cooking,' a serious game for mild cognitive impairment and Alzheimer's disease: a pilot study". *Frontiers in Aging Neuroscience, Vol. 7*.
- [96] Quick, V., Corda, K., Chamberlin, B., et al., (2013). "Ninja kitchen to the rescue: evaluation of a food safety education game for middle school youth". *British Food Journal, Vol.115, Issue. 5*, pp. 686-699.
- [97] Ricci, K.E., (1994). "The use of computer-based videogames in knowledge acquisition and retention". *Journal of Interactive Instruction Development, Vol. 7, Issue. 1*, pp. 17–22.
- [98] Sweetser, P. and Wyeth, P., (2005). "GameFlow: a Model for Evaluating Player Enjoyment in Games". *Computers in Entertainment, Vol.3, Issue. 3*.
- [99] Hsu, C. L. and Lu, H. P., (2004). "Why do People Play On-Line Games? An Extended TAM with Social Influences and Flow Experience". *Information and Management, vol.41, Issue. 7*.
- [100] Brom, C., Sisler, V. and Slavik, R., (2010). "Implementing Digital Game-Based Learning in Schools: Augmented Learning Environment of 'Europe 2045'". *Multimedia Systems, Vol. 16, No. 1*, pp.23-41.
- [101] Yusoff, A., Crowder, R. and Gilbert, L., (2010). "Validation of Serious Games Attributes Using the Technology Acceptance Model". In *2010 Second International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)*.
- [102] Eichenberg, C., Grabmayer, G. and Green, N., (2016). "Acceptance of Serious Games in Psychotherapy: An Inquiry into the Stance of Therapists and Patients". *Telemed J E Health, Vol. 22*, pp. 945-951.
- [103] Enah, C., Piper, K. and Moneyham, L., (2014). "Qualitative evaluation of the relevance and acceptability of a web-based HIV prevention game for rural adolescents". *Journal of Pediatric Nursing, Vol. 30, Issue. 2*, pp. 321-328.
- [104] Lanyi, S., Brown, J., Standen, P. et al., (2012). "Results of user interface evaluation of serious games for students with intellectual disability". *Acta Polytechnica Hungarica, Vol. 9*, pp. 225-245.
- [105] Mikovec, Z., Salvik, P. and Zara, J., (2009). "Cultural Heritage, User Interfaces and Serious Games at CTU Prague". *15th International Conference on Virtual Systems and Multimedia*, pp. 211-116.
- [106] Bonakdarian, E. and White, L., (2004). "Robocode throughout the curriculum". *Journal of Computing Sciences in Colleges, Vol. 19, Issue. 3*, pp. 311-313.